Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.2, April 2025, pp. 364~372
e-ISSN: 2548-6861

Comparing Decision Tree and Support Vector Machines in Hospital Satisfaction

Dinda Anggraini [1]*, Indra Gamayanto [2]*, Sasono Wibowo [3]*
* Sistem Informasi, Universitas Dian Nuswantoro
dindaagr25@gmail.com [1], indra.gamayanto@dsn.dinus.ac.id [2], sasono.wibowo@dsn.dinus.ac.id [3]

Article Info

Article history:

Received 2025-02-25
Revised 2025-03-04
Accepted 2025-03-07

Keyword:

Decision Tree,
SVM,
Visitor Satisfaction,
Classification,
Hospital Service Optimization.

ABSTRACT

Patient satisfaction is a key indicator of hospital service quality. This study compares the performance of Decision Tree and Support Vector Machine (SVM) in classifying patient satisfaction at Harapan Hospital Magelang for service optimization. The dataset, derived from a 2024 survey, consists of 577 samples and 13 predictor variables, covering patient demographics and medical service aspects. Preprocessing includes data cleaning, normalization, encoding, and class balancing using SMOTE. The Decision Tree is applied with gini impurity and max_depth=11, while SVM uses the RBF kernel (C=100, gamma=0.01). Model evaluation metrics include accuracy, precision, recall, F1-score, and ROC-AUC.Results show that Decision Tree outperforms SVM, achieving 86% accuracy vs. 81%. It also has 86% precision and 95% recall for the Dissatisfied category, higher than SVM (93% recall). The McNemar test confirms a statistically significant performance difference (p-value = 0.037). With higher accuracy and interpretability, Decision Tree is recommended as the primary method for hospital patient satisfaction analysis. These findings support the development of an adaptive classification system for Indonesian healthcare data.

I. Introduction

Patient satisfaction is a crucial parameter in assessing the quality of health services and is a determining factor in the sustainability of hospital operations in this competitive era. Harapan Magelang Hospital, one of the leading health service institutions in the Magelang region, faces significant challenges in systematically and measurably analyzing and improving patient satisfaction systematically and measurably. The level of patient satisfaction significantly correlates with the quality of health services, which directly impacts the reputation and sustainability of hospital operations[1]. Patient satisfaction analysis is increasingly important because competition between health service providers is getting tighter, and people increasingly demand quality health services. The development of machine learning technology, especially the Decision Tree and Support Vector Machine (SVM) methods, opens up new opportunities in classifying and analyzing patient satisfaction more accurately and efficiently, increasing prediction accuracy by up to 85% compared to conventional methods[2]. Manual analysis cannot accommodate the data's complexity, which includes various variables such as wait times, duration of treatment, and patient feedback. Decision Trees and SVM offer different approaches to classification. Decision Trees excel in outcome interpretability and decision-making transparency, while SVM has superior capabilities in handling non-linear and complex data. Applying these two methods is expected to help identify factors that affect the decline in the quality of inpatient services and increase the efficiency of outpatient service capacity.

Previous research revealed that hospitals that implemented machine learning-based data analysis experienced a 30% increase in operational efficiency and a 25% increase in patient satisfaction[3]. That emphasizes the importance of selecting the proper classification method to optimize patient satisfaction analysis. However, there has been no comprehensive study comparing the effectiveness of Decision Tree and SVM in the context of hospital patient satisfaction in Indonesia, especially with data characteristics such as at Harapan Magelang Hospital, and there is no consensus on the most effective classification method. Other research conducted in selecting the proper classification method shows that it can increase the accuracy of patient satisfaction prediction by up to 90% and assist hospital management in making strategic decisions for service improvement[4].

The integration of machine learning-based patient satisfaction analysis systems requires an in-depth understanding of the data's characteristics and adequate infrastructure. Harapan Magelang Hospital, with a dynamic pattern of visits, as shown by the data for the period November 2023 to October 2024, requires a robust analysis method to accommodate the complexity of the existing data. The successful implementation of the patient satisfaction classification system is highly dependent on the suitability of the chosen method with the characteristics of the institution's specific data.

This study aims to examine and compare the accuracy level between the Decision Tree method and the Support Vector Machine (SVM) in classifying the level of patient satisfaction, as well as identify the factors that affect the performance of the two methods in the context of hospital services. This study also aims to formulate strategic recommendations for service optimization based on the results of classification analysis. Machine learning methods, especially Decision Tree and Support Vector Machine (SVM), offer a systematic and measurable approach to classifying and understanding patient satisfaction levels. Decision Tree was chosen because of its ability to identify key factors that affect patient satisfaction in an easy-to-understand manner. This method is also flexible in handling data with non-linear patterns. It is suitable for analyzing various variables contributing to the patient experience, such as treatment duration, wait time, and patient feedback.

Meanwhile, SVM has the advantage of handling high-dimensional data and is able to find the optimal dividing line to distinguish customer satisfaction categories more accurately. Although other methods, such as Random Forest and Neural Networks, have the potential to improve classification accuracy, the complexity of the model is often an obstacle in implementation in hospital settings, which requires high interpretability and efficiency in decision-making. Ensemble models like Random Forest can indeed reduce the risk of overfitting, but the study focuses not only on improving accuracy, but also on understanding customer satisfaction patterns more deeply. In addition, methods such as Naïve Bayes are less suitable due to the assumption of independence between variables that are rarely found in patient satisfaction data, while Logistic Regression tends to be less effective in handling non-linear relationships in customer satisfaction analysis.

In the context of Harapan Magelang Hospital, which has a dynamic visit pattern based on data from November 2023 to October 2024, the selection of the right classification method is a determining factor in understanding patient satisfaction trends more accurately. Therefore, this study will evaluate and compare the performance of Decision Tree and SVM in classifying patient satisfaction, as well as identify key variables that affect patient satisfaction levels. In addition, the results of this study will also be used to formulate strategic recommendations to optimize hospital services based on the results of classification analysis.

The dataset used in this study comes from the results of the patient satisfaction survey of Harapan Magelang Hospital conducted in 2024, with a total of 577 data consisting of 13 predictor variables. These variables include name, gender, age, last education, worker, frequency of visits, type of payment, and aspects of medical services. Aspects of medical services include service responsiveness (waiting time, emergency measures, satisfaction of complaints, patient registration), empathy (staff attention, communication skills, coordination between staff, explanation of medical procedures), cleanliness of the place (cleanliness of the room, bathroom facilities, waiting room, accessibility of facilities), certainty (expertise of medical staff, environmental safety, clarity of cost information, suitability of costs with services), reliability (satisfaction with treatment results, arrangements schedule, additional cost transparency, insurance information), and targets. Comparative analysis will utilize standard evaluation metrics, such as accuracy, precision, recall, and F1-score. In addition, aspects of interpretability and computational complexity will also be considered in this study

The main contributions of this study include the development of a comparative framework for the evaluation of patient satisfaction classification methods specific to the context of hospitals in Indonesia, the identification of key factors influencing the effectiveness of Decision Tree and SVM methods in patient satisfaction classification, and the formulation of strategic recommendations for service optimization based on classification results. An in-depth understanding of the performance of classification methods can assist health institutions in developing more effective service improvement strategies. This research is expected to contribute to the development of the internal analysis system of Harapan Magelang Hospital and be a reference for other health institutions that are implementing similar systems.

The significance of this study also lies in its potential to develop a classification model that is adaptive to the unique characteristics of Indonesian health data. That is in line with the findings of recent research that considers the interpretability aspect of classification results to ensure that the analysis results can be translated into actual actions in service improvement[5]. Efforts to improve the quality of health services through systematic patient satisfaction analysis are strategic actions that must be taken to face the increasingly fierce competition in the health sector. The ability to effectively analyze and respond to patient feedback will be a key factor in maintaining and improving the hospital's competitive position in the future. Harapan Magelang Hospital can develop a more measurable and effective service improvement strategy by understanding patient satisfaction patterns and implementing appropriate classification methods.

II. METHOD

This study applies a quantitative approach by comparing the Decision Tree method and the Support Vector Machine (SVM) to classify the level of patient satisfaction at Harapan Magelang Hospital. Based on data obtained from November 2023 to October 2024, 9,830 inpatients and 149,018 outpatients were recorded, with an increasing trend in the average outpatient visits but accompanied by a decrease in the average inpatients.

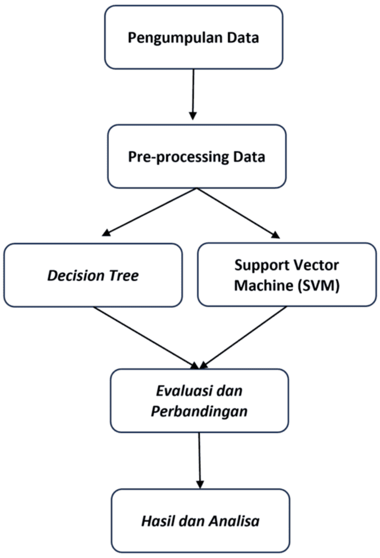

Figure 1. Research flow chart

Figure 1 shows the flow of research methodology implemented in the comparative analysis of Decision Tree and SVM methods for classifying patient satisfaction at Harapan Magelang Hospital. The methodology of this research includes several stages that are carried out systematically and structured. The first stage begins with collecting patient data, including data on inpatient and outpatient visits and various parameters related to patient satisfaction. The collected data enters the preprocessing stage to ensure data quality and consistency. After the preprocessing process, the data is separated into two parts with a ratio of 90:10, namely 90% for training and 10% for testing.

Furthermore, data analysis was carried out in parallel using two classification methods: Decision Tree and Support Vector Machine (SVM). These two methods are chosen based on complementary characteristics, where the Decision Tree has the advantage of interpretability of results. In contrast, SVM has a good ability to handle non-linear data. The next stage is to explore the model's performance using test data. In this phase, the built model predicts the test data. The prediction results are then compared with the actual values in the test data to measure the model's error level. Model performance evaluation uses measurement metrics such as accuracy, precision, and recall. The final stage of this research methodology is service optimization. At this stage, the analysis and comparison of the two methods are used as the basis for developing recommendations for improving the quality of hospital services. Each stage in this methodology is systematically designed to produce accurate analysis and can be implemented practically in hospital operations.

A. Data Collection

According to Kristiawan and Widjaja, comprehensive data collection is a critical factor in the analysis of hospital visitor satisfaction[6]. The research data was obtained based on a survey that made a questionnaire available to outpatient and inpatient polyclinic patients at Harapan Magelang Hospital. These data were collected during the study period from hospital patients' respondents. In this study, the Likert scale is used as a measurement tool for the data. A person's response to an event or social phenomenon can be measured using the Likert Scale as one of the measurement methods in research[7]. The answers in the questionnaire are in the form of a choice from the existing criteria values, namely:

1: Very Dissatisfied
2: Dissatisfied
3: Quite Satisfied
4: Satisfied
5: Very Satisfied

TABLE I
CLASSIFICATION OF ASSESSMENT ATTRIBUTES

Questionnaire Code	**A1**	**A2**	**A3**	**A4**
Service Response (P1)	Action Wait Time	Emergency Measures	Satisfaction of Complaint Submission	Patient Registration
Questionnaire Code	A5	A6	A7	A8
Empathy (P2)	Staff Attention	Communication Skills	Medical Procedure Explained	Coordination Between Staff
Questionnaire Code	A9	A10	A11	A12
Cleanliness of the Place (P3)	Room Cleanliness	Bathroom Facilities	Lounge	Accessibility of Facilities
Questionnaire Code	A13	A14	A15	A16
Certainty (P4)	Medical Staff Expertise	Environmental Safety	Clarity of Fee Information	Cost Compatibility with Services
Questionnaire Code	A17	A18	A19	A20
Reliability (P5)	Satisfaction with Treatment Results	Schedule Setting	Transparency of Additional Costs	Information about Insurance

This study utilizes a dataset that is divided into two types of attributes. The first group includes seven essential attributes, namely gender, age, last education, occupation, frequency of visits, type of service payment, and satisfaction label. The second group consists of five main attributes related to hospital services: service responsiveness, empathy, cleanliness of the place, certainty, and reliability. This dataset allows for a more detailed and accurate classification of patient satisfaction using the Decision Tree and SVM methods. Each of these five key attributes has four more specific assessment sub-attributes, as shown in Table I.

B. *Preprocessing Data*

Preprocessing is an important step in the data mining stage[8]. Preprocessing aims to clean, standardize, and prepare data so that it can be processed more efficiently by analysis algorithms or models of Machine Learning[9]. Process Preprocessing the data is carried out to ensure that the data to be analyzed is of good quality. Referring to the research conducted by Siallagan et al., the preprocessing data includes [10]:

1) Data cleaning to handle missing values and remove duplicate data
2) Normalization of numerical data for min-max scale for SVM because it is sensitive to data scale
3) Categorical data encoding using one-hot encoding
4) Feature selection based on relevance to patient satisfaction

The dataset that has gone through the preprocessing stage is then the used as input in the classification model to compare the Decision Tree and SVM methods. Adjustments to the data scale are made to improve the accuracy of the model and avoid bias due to differences in the value range between attributes. In this study, the dataset used amounted to 577 data with two target classes: dissatisfied and satisfied. The distribution of data can be seen in the figure below.

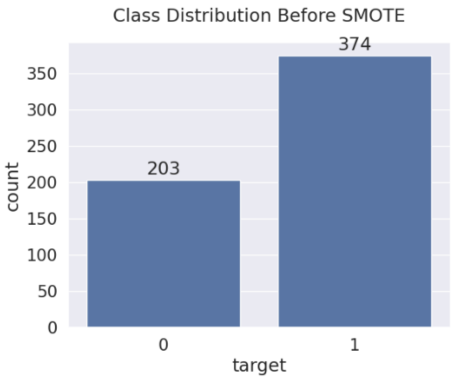

Figure 2. Class distribution before SMOTE

The figure above shows 203 satisfied patient data and 374 dissatisfied patient data. This shows a class imbalance. Therefore, SMOTE (Synthetic Minority Over-Sampling Technique) is applied to balance the data in each class.

In the figure above, after the SMOTE process is carried out, 374 patient data are satisfied, and 374 patient data are dissatisfied. Thus, after SMOTE is implemented, the data becomes balanced in each class.

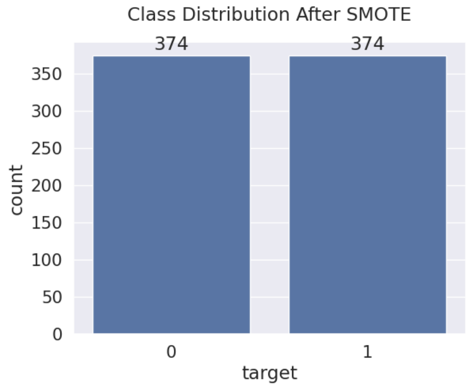

Figure 3. Class distribution after SMOTE

C. *Decision Tree*

A decision tree structure to describe and determine decisions based on several predetermined rules and conditions[11]. This machine learning method can be used for classification, i.e., predicting the target class from the data provided. Decision trees work by classifying them based on a hierarchical set of decision rules and forming a tree structure, where each leaf node in the tree represents a decision based on data features[12]. This decision tree structure will be used to make decisions based on the features that exist in the data. In building a decision tree, the attribute with the highest information gain value is selected as the root to separate the data classes in the most effective way[13]. The formula for impurity is as follows.

$$Entropy(S) = \sum_{i=0}^{n} -p_i * log^2(p_i) \quad (1)$$

Information:
S : Case set
N : Number of partitions S
Pi : Number of cases in partition – I

It then calculates the Gain value to measure the decrease in uncertainty after data sharing based on certain attributes.

$$Gain(S,A) = Entropy(S) - \frac{\sum_{i=1}^{n}|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Information:
S : Case set
A : Number of Partitions attribute A
|Ni| : Number of cases in partition – I
|N| : Number of cases and N

D. *Support Vector Machine*

Support Vector Machine (SVM) is one of the supervised machine learning models used to determine classification or regression[14]. The Support Vector Machine (SVM) operates by finding the best hyperplane capable of separating two classes in the input space by maximizing the margin between the two classes[15]. This method efficiently manages On data that cannot be separated linearly, SVM utilizes kernel tricks to map data to higher-dimensional feature spaces[16]. The separator hyperplane in SVM can be mathematically expressed as wx+b=0, where w is the weight vector, x is the input vector, and b is the bias [17][18]. Data needs to be transformed from a non-linear form to a linear form through a process known as kernelization The Support Vector Machine (SVM) has different kernels, including linear kernels, polynomial kernels, Gaussian Radial Basis Function (RBF) kernels, and sigmoid kernels. Here are the similarities that each type of kernel has[20].

1. *Linear Kernel*

$$K(x_i, x) = x_i^T x \tag{3}$$

Information:
x_i : training data
x : testing data

2. *RBF Kernel*

$$K(x_i, x) = exp\ exp\left(-\frac{|x_i - x|^2}{\sigma}(2\sigma^2)\right) \tag{4}$$

Information:
x_i : training data
x : testing data
σ : scale parameter

3. *Polynomial Kernel*

$$K(x_i, x) = (\gamma(x_i^T x) + r)^p \tag{5}$$

Information:
x_i : training data
x : testing data
p : polynomial degree

4. *Sigmoid Kernel*

$$K(x_i, x) = tanh(\gamma(x_i^T x) + r) \tag{6}$$

Information:
x_i : training data
x : testing data
r : coefficient

E. Evaluation

The classification model will be evaluated using the Confusion Matrix, which consists of precision, recall, accuracy, and f1-score measurements. Metric accuracy, precision, recall, and F1-score were chosen because each provides a different perspective in evaluating the performance of the classification model, especially in the context of class imbalances that often occur in hospital customer satisfaction data. Accuracy Measure the proportion of correct predictions, but not enough if there is an inequality in class distribution. Precision Assessing the extent to which optimistic predictions are relevant is important in preventing misclassification of high satisfaction that can mislead hospital management. Recall that measuring the model's ability to detect all specific satisfaction cases is crucial to understanding the level of satisfaction as a whole. F1-score Combine precision and recall in one balanced metric, it provides a more comprehensive picture of the model's performance, especially when one metric is not sufficiently representative of the overall model performance. This test is important to ensure the quality of the model that has been created and to find out how well the system performs in classifying data. Rahmini Hadi et al. recommend the use of a confusion matrix for a detailed analysis of classification performance, focusing on[21]:

1. True Positives (TP)
2. True Negatives (TN)
3. False Positives (FP)
4. False Negatives (FN)

F. Service Optimation

The analysis's results are used to improve hospital services by identifying areas for improvement, developing data-based recommendations, implementing changes systematically, and conducting continuous monitoring and evaluation.

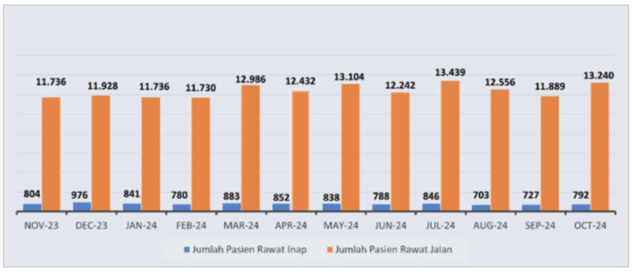

Figure 4. Number of inpatients and outpatients in november 2023—october 2024

Based on the data in the graph above, a significant pattern can be observed in the distribution of patient visits at Harapan Magelang Hospital from November 2023 to October 2024. The total number of inpatients was recorded at 9,830, while the total number of outpatients reached 149,018. This data shows the dominance of outpatient services in hospital operations. Trend analysis shows contrasting patterns between inpatient and outpatient services. The number of outpatient visits consistently increased from around 12,500 visits per month in November 2023 to 14,800 visits in October 2024, showing a growth of around 18.4%.

On the other hand, inpatient services experienced a gradual decline from 850 patients per month to 730 patients, indicating a decrease of 14.1%. This pattern is an important basis for applying the Decision Tree and SVM methods to analyze the factors that affect patient satisfaction. This historical data will be used as a training dataset for both methods, considering variables such as wait time, duration of treatment, and patient feedback. A comparison of the two methods will help identify the most effective approach to optimizing hospital services based on the identified visit patterns. This pattern indicates several important aspects to consider in the analysis:

1) Shift in patient preferences from inpatient to outpatient
2) Potential areas of improvement in inpatient services
3) The need to optimize outpatient service capacity
4) Evaluation of the effectiveness of the referral system

5) Analysis of factors influencing the decrease in hospitalization

This data will be a key component in the preprocessing stage and the formation of classification models using Decision Trees and SVMs to generate the proper recommendations for optimizing hospital services.

III. RESULT AND DISCUSSION

A. Preprocessing Data

Before the dataset is processed using a predefined algorithm, it will be processed as the Preprocessing data. The dataset of 577 data with 13 attributes will be further processed at this stage. In this process, stages such as cleaning data from duplicate data and missing data and transforming data from categorical to numerical are carried out so that it can be processed more easily later and applied to both algorithms. The data transformation process is carried out for type and labels. That type is based on gender, age, last education, occupation, visit frequency, and service payment. As for the labels, they are taken from the scale data Likert. For this scale, if the average scale from the sum of P1 to P5 < 3 = 0, and if the average scale of the sum of P1 to P5 the scale ≥ 3 = 1. Data processed after the Preprocessing is as follows in table II.

TABLE II
DATASET AFTER PREPROCESSING

Gender	Age	Education	Work	Frequency	Payment Type	P1	P2	P3	P4	P5	Target
1	1	5	2	1	4	3	3,75	3,5	3,75	3,5	1
0	2	6	4	3	1	3	3,25	3	2,75	3,5	1
1	1	3	2	1	4	2,75	3,5	3,75	4	3	1
0	1	4	2	2	1	2,75	3,25	3	3,25	3	1
1	1	3	2	2	1	3,25	3,5	3,75	3,5	3,25	1
1	1	3	4	2	2	3	3	3	3,25	3	1
1	1	3	4	2	2	2,75	2	3,25	3,5	3,25	1
0	1	3	4	1	1	3,75	4	4	3,5	3,5	1
1	2	5	6	1	1	3,25	3,5	3	3,75	3,5	1
1	1	5	6	3	1	3,25	3,25	2,75	3,25	3	1

B. Data Processing

After the preprocessing process, the next step is the data processing stage. The data management process is carried out at this stage by applying the Decision Tree algorithm and the Support Vector Machine (SVM). The data to be processed is divided into two parts, namely training and test data, before applying the algorithm. This data is separated by the sets' percentages of 90 and 10%, respectively.

1) Decision Tree Classification

After The data that has been divided will be carried out to the next stage, namely the application of the Decision Tree algorithm to check the evaluation of the classification method by building with criterion="gini", max_depth="11", min_samples_leaf="7", and min_samples_split="2". Where the results of these parameters are obtained through the hyperparameter tuning process using grid search with an iteration value of 3 to find the best parameter parameters used, the parameters used in the Decision Tree, namely criterion="gini", max_depth=11, min_samples_leaf=7, and min_samples_split=2, influence model performance, especially in terms of accuracy and risk of overfitting. The "gini" criterion is used to measure the degree of imbalance of a node, so the model chooses a separation that minimizes impurity. Max_depth=11 limits the depth of the decision tree to 11 levels, which aims to reduce model complexity and prevent overly detailed training on training data. Min_samples_leaf=7 ensures that each leaf node should have a minimum of 7 samples, which helps to avoid excessive data fragmentation.

Meanwhile, min_samples_split=2 allows each node to split if it has at least two samples, making the tree too complex if not combined with other control strategies such as pruning or cross-validation. Since Decision Trees naturally tend to overfit training data, this model should be supported by pruning techniques (post-pruning or pre-pruning) and cross-validation to improve generalization in test data. In this test, we use parameter selection with Grid Search for model optimization with cv="3."

Decision Trees have several advantages in the context of hospital customer satisfaction analysis. The model is easy for hospital managers to understand and interpret because the results can be visualized as a clear tree diagram showing the key factors that affect patient satisfaction. In addition, Decision Tree can handle both categorical and numerical data without requiring much preprocessing. However, the downside is that the model is prone to overfitting, especially if the data has many variables with few observations. Additionally, Decision Tree models can be biased if there is a class imbalance in the data, leading to less accurate decisions against certain groups. The information on the application results uses the Decision Tree algorithm in table III.

TABLE III
DECISION TREE ALGORITHM PERFORMANCE

***Accuracy*: 86%**			
	***True* Satisfied**	***True* Not Satisfied**	***Class Precision***
Pred Satisfied	12	2	86%
Dissatisfied Pred	6	38	86%
Class Recall	67%	95%	

Based on the evaluation results shown in table III, the implementation of the Decision Tree method produces performance with an overall accuracy level of 86%. Regarding model sensitivity, a % recall value of 67% was obtained for the Satisfied category and 95% for the Dissatisfied category. Meanwhile, the level of accuracy of predictions measured through precision scores reached 86% for the Satisfied classification and 86% for the Dissatisfied classification. This performance reflects several classification errors in the system, where 6 cases of dissatisfaction are incorrectly categorized as Satisfied and 12 cases of satisfaction are wrongly classified as Dissatisfied. A visualization of the performance metrics of the Decision Tree algorithm can be observed in the figure III.

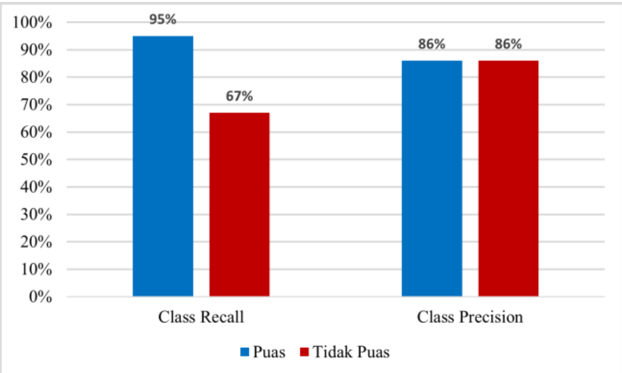

Figure 5. Visualization of decision tree algorithm performance

2) *Support Machine Vector Classification*

Support Vector Machine (SVM) method is almost identical to the Decision Tree method, but the difference lies only in the type of parameter used. This Support Vector Machine (SVM) method was developed by utilizing kernel type = "rbf", C="100", and gamma = "0.01" obtained by the hyperparameter tuning process using Grid Search with an iteration value of 3 to find the best parameter used.

The parameters used in the Support Vector Machine (SVM) significantly influence model performance, particularly affecting accuracy, model complexity, and the ability to generalize to new data. The RBF (Radial Base Function) kernel was chosen because it can handle linearly inseparable data, improving the model's performance in detecting complex patterns in the hospital customer satisfaction dataset. The parameter C=100 serves as a regulatory factor, where a more significant value tends to make the model more focused on the training data and potentially leads to overfitting. Meanwhile, gamma=0.01 determines how far the influence of one data point can reach another point in the feature space; The smaller the gamma value, the wider the influence on the distribution of data.

Because SVM is sensitive to data scale, the normalization process is performed on all numerical variables to ensure that the model can perform optimally without bias due to scale differences between features. To improve the generalization of the model to the test data, cross-validation with cv=3 is used in parameter selection.

In the context of hospital customer satisfaction analysis, SVM has several advantages. The model effectively classifies high-dimensional data and can handle non-linear relationships between variables. If the parameters selected are proper, SVMs are more resistant to overfitting than Decision Trees. However, the disadvantage is that interpreting SVM results is more complex, especially when explaining the main factors that influence the model's decisions. In addition, SVM compute times are longer than Decision Trees, especially on large datasets.

Information on the results of the application with the Support Vector Machine (SVM) method in table IV.

TABLE IV
SUPPORT VECTOR MACHINE ALGORITHM PERFORMANCE

***Accuracy*: 81%**			
	***True* Satisfied**	***True* Not Satisfied**	***Class Precision***
Pred Satisfied	10	3	77%
Dissatisfied Pred	8	37	82%
Class Recall	56%	93%	

Referring to the test results presented in table IV, the application of the Support Vector Machine (SVM) algorithm shows an achievement with an overall accuracy of 81%. Regarding the sensitivity level of the model, a recall value of 56% was obtained in the Satisfied category and 93% in the Dissatisfied category. On the other hand, the level of prediction accuracy reflected through the precision value shows a figure of 77% for the Satisfied classification and 82% for the Dissatisfied classification.

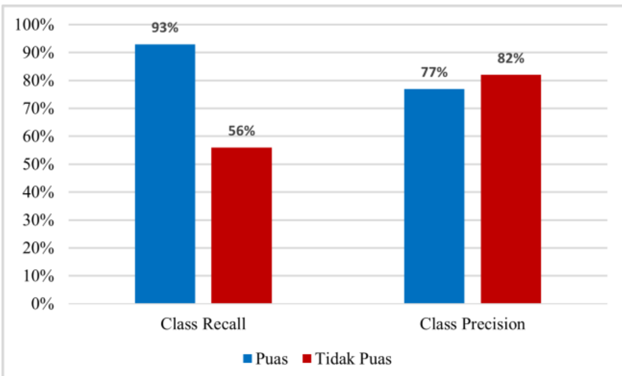

Figure 6. Visualization of the Performance of the Support Vector Machine (SVM) algorithm

This performance indicates several classification inaccuracies in the system, namely 8 cases of dissatisfaction incorrectly identified as Satisfied and 3 cases incorrectly categorized as Dissatisfied. A graphical representation of the performance metrics of the Support Vector Machine (SVM) algorithm can be observed in image IV.

3) Evaluation and Comparison of Decision Tree and Support Vector Machine (SVM) Classification

Based on the data processing stages that have been carried out previously, the comparison of the performance between the two methods can be seen in table V.

TABLE V
COMPARISON OF PERFORMANCE ALGORITHM

Algorithm	**Accuracy**	***Recall***		**Precision**	
		***True* Satisfied**	***True* Not Satisfied**	**Pred Satisfied**	**Dissatisfied Pred**
Decision Tree	86%	67%	95%	86%	86%
SVM	81%	56%	93%	77%	82%

Based on the results of comparative testing of the performance of the classification algorithm, it was obtained that the Decision Tree algorithm obtained a more optimal overall accuracy level of 86% compared to the Support Vector Machine (SVM) algorithm, which obtained an accuracy of 81%. Analysis of the recall value in the 'Satisfied' category shows that Decision Tree has a higher sensitivity with a value of 67%, while SVM obtained a score of 56%.

An evaluation of the 'Dissatisfied' category indicates that both algorithms have excellent detection capabilities, with Decision Tree achieving a recall value of 95% and SVM reaching 93%. In terms of precision, the Decision Tree algorithm demonstrated consistency with a score of 86% for both classification categories. In comparison, SVM produced varying values of 77% for the "Satisfied" category and 82% for the "Dissatisfied" category.

A statistical test, the McNemar test, is carried out to determine whether the difference is significant, considering that the data comes from binary classification.

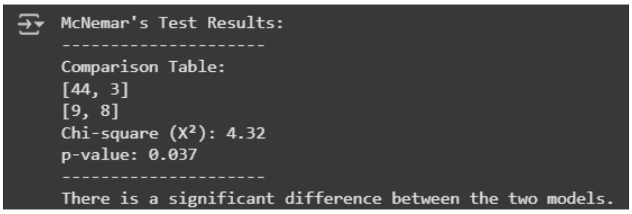

Figure 7. Result of McNemar test

From the test results table, the number of instances classified differently by the two models is calculated. Based on the calculations, the statistical value of the McNemar test was obtained as $X^2 = 4.32$ with p-value = 0.037. Since the p-value < 0.05, it can be concluded that the difference in accuracy between the Decision Tree and the SVM is statistically significant at a 95% confidence level. In addition to accuracy, Decision Tree also showed higher recall than SVM, especially in the "True Not Satisfied" class (67% vs. 56%). This shows that the Decision Tree better detects "Dissatisfied" cases than SVMs. The Decision Tree precision for "Pred Satisfied" is also higher (95% vs. 93%), indicating that the model is more reliable in classifying customer satisfaction. The test and statistical analysis results show that Decision Tree performs superior to SVM, with statistically significant differences.

The evaluation results show that implementing the Decision Tree algorithm provides superior performance in overall accuracy, consistency of precision values, and the ability to identify "Satisfied" cases. Nevertheless, both algorithms detected "Dissatisfied" cases with recall values exceeding 90%. These findings can be the basis for consideration when selecting classification algorithms that suit the system's needs. The decision tree algorithm is proven to have the most optimal accuracy. However, it can still be improved through ensemble learning in the training process so that the accuracy can be more optimal, such as using random forests or combinations with other algorithms.

IV. CONCLUSION

Based on the findings of the research that has been conducted regarding the application and comparison of the Decision Tree algorithm and the Support Vector Machine (SVM) for customer satisfaction classification, it is obtained that the Decision Tree algorithm has superior performance with a global accuracy of 86% compared to SVM which reaches 81%. In addition, the Decision Tree showed a more consistent precision of 86% for both categories (Satisfied and Dissatisfied). At the same time, SVM had varying precision, namely 77% for the Satisfied category and 82% for the Dissatisfied category. Regarding recalls, Decision Tree also excels at identifying dissatisfied customers, with 95% recalls compared to 93% on SVMs.

The McNemar statistical test applied in this study showed that the difference in accuracy between Decision Tree and SVM was statistically significant with p-value = 0.037, which confirmed that Decision Tree was more effective in classifying hospital patient satisfaction. In addition to the advantages in accuracy, the Decision Tree is also easier to interpret than SVM because its classification results can be visualized in the form of a clear decision tree. That makes it easier for hospital management to understand and apply in analyzing patient satisfaction and developing service improvement strategies.

Although the Decision Tree shows better results, this model has the potential for overfitting, especially when the tree depth is too large. Therefore, applying pruning and cross-validation techniques is recommended to improve the model's generalization against the new data. Overall, this study shows that Decision trees are more effective in analyzing hospital patient satisfaction than SVM, both in accuracy, interpretability, and reliability in detecting patient dissatisfaction.

For further research development, it is recommended that more sophisticated data processing techniques be applied and that merger methods be explored to improve classification accuracy. System development can also be directed towards direct implementation integrated with customer relationship management systems to enable faster follow-up of identified dissatisfied customers.

REFERENCES

[1] H. Wijaya, A. Rohendi, K. Mulyani, and U. A. R. Sanjaya, “Pengaruh Kepercayaan, Kualitas Pelayanan, dan Kewajaran Harga Terhadap Kepuasan Pasien Di Klinik S, Tangerang Selatan”.

[2] M. I. Sodikin, “Penerapan dan Manfaat Machine Learning di Rumah Sakit,” Multiverse Open Multidiscip. J., vol. 2, no. 2, pp. 262–265, Aug. 2023, doi: 10.57251/multiverse.v2i2.1207.

[3] N. Lelyana, “Analisis Dampak Inovasi Teknologi pada Strategi Manajemen Rumah Sakit,” JISHUM J. Ilmu Sos. Dan Hum., vol. 2, no. 4, pp. 425–446, Jun. 2024, doi: 10.57248/jishum.v2i4.380.

[4] R. Hamdani and A. Darta, “Analisis Tingkat Kepuasan Pasien Di Klinik Pratama Salbiyana Dengan Algoritma C4.5,” J. Sci. Soc. Res., vol. VII, pp. 273–280, 02/24.

[5] A. Gunawan and S. Kom, “Pengantar Sistem Informasi Kesehatan”.

[6] K. Kristiawan and A. Widjaja, “Perbandingan Algoritma Machine Learning dalam Menilai Sebuah Lokasi Toko Ritel,” J. Tek. Inform. Dan Sist. Inf., vol. 7, no. 1, Apr. 2021, doi: 10.28932/jutisi.v7i1.3182.

[7] A. Ulfah, D. Hermina, and N. Huda, “Desain Instrumen Evaluasi Yang Valid Dan Reliabel Dalam Pendidikan Islam Menggunakan Skala Likert,” J. Studi Multidisipliner, vol. 8, no. 12, pp. 855–861, Dec. 2024.

[8] M. D. Purbolaksono, M. Irvan Tantowi, A. Imam Hidayat, and A. Adiwijaya, “Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes,” J. RESTI Rekayasa Sist. Dan Teknol. Inf., vol. 5, no. 2, pp. 393–399, Apr. 2021, doi: 10.29207/resti.v5i2.3008.

[9] F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmaddeni, and L. Efrizoni, “Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia: Application of K-Nearest Neighbor Algorithm Using Wrapper as Preprocessing for Determination of Human Weight Information,” Malcom Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 1, pp. 273–281, Jan. 2024, doi: 10.57152/malcom.v4i1.1085.

[10] E. Siallagan, I. Parlina, and D. Suhendro, “Model Aturan Tingkat Kepuasan Pasien Terhadap Pelayanan Puskesmas Menggunakan Algoritma C4.5,” Data Sci., vol. 1, no. 2, 2022.

[11] D. A. Mukhsinin, M. Rafliansyah, S. A. Ibrahim, R. Rahmaddeni, and D. Wulandari, “Implementasi Algoritma Decision Tree untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix: Implementation of Decision Tree Algorithm for Movie Recommendation and Rating Classification on the Netflix Platform,” MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 2, pp. 570–579, Mar. 2024, doi: 10.57152/malcom.v4i2.1255.

[12] M. Bansal, A. Goyal, and A. Choudhary, “A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning,” Decis. Anal. J., vol. 3, p. 100071, Jun. 2022, doi: 10.1016/j.dajour.2022.100071.

[13] I. Arfyanti, M. Fahmi, and P. Adytia, “Penerapan Algoritma Decision Tree Untuk Penentuan Pola Penerima Beasiswa KIP Kuliah,” Build. Inform. Technol. Sci. BITS, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2275.

[14] J. Kusuma, A. Jinan, M. Z. Lubis, R. Rubianto, and R. Rosnelly, “Komparasi Algoritma Support Vector Machine Dan Naive Bayes Pada Klasifikasi Ras Kucing,” Generic, vol. 14, no. 1, pp. 8–12, Jan. 2022, doi: 10.18495/generic.v14i1.122.

[15] K. A. Rokhman, B. Berlilana, and P. Arsi, “Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online,” J. Inf. Syst. Manag. JOISM, vol. 3, no. 1, pp. 1–7, Jan. 2021, doi: 10.24076/JOISM.2021v3i1.341.

[16] A. M. Yolanda and R. T. Mulya, “Implementasi Metode Support Vector Machine untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store”.

[17] A. Sudin, M. Salmin, M. Fhadli, and A. M. Mamonto, “Klasifikasi Kelayakan Air Minum Bagi Tubuh Manusia Menggunakan Metode Support Vektor Machine Dengan Backward Elimination”.

[18] M. Rafly and I. Veritawati, “Penggunaan Algoritma Support Vector Machine Untuk Penentuan Rekomendasi Penerima Beasiswa Di SMA Negeri 8 Kota Bogor,” IKRA-ITH Inform. J. Komput. Dan Inform., vol. 9, no. 1, pp. 102–113, Oct. 2024, doi: 10.37817/ikraith-informatika.v9i1.4381.

[19] R. Oktafiani and R. Rianto, “Perbandingan Algoritma Support Vector Machine (SVM) dan Decision Tree untuk Sistem Rekomendasi Tempat Wisata,” J. Nas. Teknol. Dan Sist. Inf., 2023.

[20] A. Z. Praghakusma and N. Charibaldi, “Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter,” JSTIE J. Sarj. Tek. Inform. E-J., vol. 9, no. 2, p. 88, Jun. 2021, doi: 10.12928/jstie.v9i2.20181.

[21] R. Hadi, “Optimalisasi Manajemen Pelayanan Gerai Sehat di Layanan Kesehatan Cuma-Cuma (LKC) Dompet Dhuafa Jawa Tengah,” Mabsya J. Manaj. Bisnis Syariah, vol. 5, no. 1, pp. 59–78, May 2023, doi: 10.24090/mabsya.v5i1.8236.